THE PYKE ESTIMATOR AND THE EXPONENTIAL DISTRIBUTION

Hugh Allan Kelley

NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

The Pyke Estimator and the Exponential Distribution

by

Hugh Allan Kelley

Thesis Advisor: R. R. Read

March 1973

The Pyke Estimator and the Exponential Distribution

by

Hugh Allan Kelley
Captain, United States Army
B.S., United States Military Academy, 1965

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 1973

ABSTRACT

It is common to estimate a distribution by means of a step function. Such estimates can be made continuous by connecting the left points of the steps with straight line segments. In this paper, the best estimator of this class is found for data which is exponentially distributed using minimum risk. The risk is then compared with those of the sample distribution function and the Pyke estimator.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF DRAWINGS

I. INTRODUCTION

The motivation behind this thesis is to provide a contribution towards the objective of finding the best way to estimate an unknown continuous distribution. The cue is taken from References 4 and 5 which have adopted the following principles:

(1) The estimator itself should be a continuous function.

(2) The estimator should be simple and natural.

The first principle has not been generally adopted. The modern textbook solution [Ref. 3] is to use the sample distribution (step) function. A simple and natural extension of the step function, satisfying both of the above principles, is to connect the left end of the steps with straight line segments, following the example of the so-called Ogive curve. Thus, the class of estimators can be described by:

(a) plot the points (C_r, X_r), $r = 1, \ldots, n$

(b) connect these points with straight line segments where $0 \leq C_1 \leq C_2 \leq \ldots \leq C_n \leq 1$ are constants determined by some rule and $X_1 < X_2 \ldots < X_n$ are the order statistics of a random sample of size n drawn from some unknown CDF, F. The only issue to be resolved is how to determine the sequence $\left\{C_i\right\}_{i=1}^{n}$.

Reference 6 suggests using $C_r = \frac{r}{n+1}$, refering to the result as the Pyke estimator, and shows that the expected squared error of a continuous Pyke estimator for some distribution, F, is no larger than the expected squared error of the sample distribution function for a sufficiently large sample size. It is also shown that the Pyke estimator strictly dominates the sample distribution function for sample sizes greater

than or equal to one. The risk function used is the integrated expected squared error, a popular choice [Ref. 1].

In Ref. 4, the sequence $\{C_i\}_{i=1}^{n}$ for which the risk is minimized is found for the case where F is the uniform distribution. Comparisons of this optimal risk with that of the sample distribution function and the Pyke estimator are made. The result is that the risk of the Pyke estimator is significantly closer to the optimal than the risk of the sample distribution function.

The purpose of this thesis is to determine if the risk of the Pyke estimator remains closer to the optimal than that of the sample distribution function for a random sample drawn from the exponential distribution. The approach will be similar to that used in Ref. 4. In an attempt to provide continuity for the reader, the notation of Ref. 4 is used wherever possible.

II. DETERMINATION OF THE MINIMUM RISK COEFFICIENTS

Let $C_0, C_1, \ldots, C_n, C_{n+1}$ be an increasing sequence with $C_0 \equiv 0$ and $C_{n+1} \equiv 1$. The continuous function estimator for F(x) can be defined by:

$$H(x) = C_r + \frac{x - X_r}{X_{r+1}-X_r}(C_{r+1} - C_r) \text{ for } X_r \le x \le x_{r+1},$$

$$r = 0, 1, \ldots, n \tag{2.1}$$

where $X_1 \le X_2 \ldots \le X_n$ are the order statistics from an absolutly continuous distribution F(x). Assume that the population is lower bounded and contained in the interval $[0,\infty)$, then define $X_0 \equiv 0$, $X_{n+1} \equiv \infty$ and the risk function by:

$$R = E \int_0^\infty \left[F(x) - H(x)\right]^2 dF(x) \tag{2.2}$$

H(x) is defined piecewise according to which of the random intervals (X_r,X_{r+1}) contains x. Assume that the sample is extracted from an exponential distribution. For convenience, let the distribution's mean equal one.[1] Let u, v with u < v be the variables in the sample space of (X_r,X_{r+1}). Their joint density function is:

$$f_{r,r+1}(u,v) = \frac{n!}{(r-1)!(n-r-1)!}(1 - e^{-u})^{r-1}e^{-u}e^{-v(n-r)}$$

$$\text{for } 0 < u < v < \infty \text{ and } 1 \le r \le n-1.$$

---

[1]The value of the mean does not matter since the risk does not change with linear transformations of the basic random variable. (Personal communication from Professor R.R. Read)

The end point densities can be derived as:

$$f_{0,1}(0,v) = ne^{-vn},\ 0 < v < \infty \tag{2.4}$$

$$f_{n,n+1}(u,\infty) = ne^{-u}(1 - e^{-u})^{n-1},\ 0 < u < \infty \tag{2.5}$$

Define $\Delta_r \equiv C_{r+1} - C_r$,

then:

$$C_r = \sum_{j=0}^{r-1} \Delta_j$$

Rewriting (2.2)

$$R = \sum_{r=0}^{n} \iiint\limits_{0 \le u \le x \le v < \infty} \left[1 - e^{-x} - C_r - \frac{x-u}{v-u}\Delta_r\right]^2 e^{-x} f_{r,r+1}(u,v)\,du\,dv\,dx \tag{2.21}$$

The minimum risk coefficients, C_r, can now be found using classical optimization techniques. Using the Lagrangain form, the problem becomes:

$$\text{Minimize: } \Phi = R - \lambda\left(\sum_{j=0}^{n} \Delta_j - 1\right)$$

$$\text{Subject to: } \sum_{j=0}^{n} \Delta_j = 1$$

where λ is the Lagrange Multiplier. Thus, the approach becomes:

Set $\frac{\partial \Phi}{\partial \Delta_k} = 0$ for $k = 0, 1, \ldots, n$ and solve for C_k.

$$\begin{aligned}\frac{\partial \Phi}{\partial \Delta_k} = &-2 \iiint\limits_{0 \le u \le x \le v < \infty} \left[1 - e^{-x} - C_k - \frac{x-u}{v-u}\Delta_k\right] \frac{x-u}{v-u} e^{-x} f_{k,k+1}(u,v)\,du\,dv\,dx \\ &-2\sum_{r=k+1}^{n} \iiint\limits_{0 \le u \le x \le v < \infty} \left[1 - e^{-x} - C_r - \frac{x-u}{v-u}\Delta_r\right] e^{-x} f_{r,r+1}(u,v)\,du\,dv\,dx \\ &+ \lambda \\ &= 0\end{aligned} \tag{2.6}$$

which after laborious but straightforward integration leads to:

$$C_k\left[2(n-k)(n-k+1) - (2n-2k+1)(n-k+1)(n-k)\ln\frac{n-k+1}{n-k}\right] - C_{k+1}$$

$$\left[(n-k)(2n-2k+1) - 2(n-k)^2(n-k+1)\ln\frac{n-k+1}{n-k}\right] + \sum_{r=k+1}^{n}$$

$$C_r\left[(n-r+1)(n-r)\ln\frac{n-r+1}{n-r} - (n-r+1)\right] - \sum_{r=K+1}^{n} C_{r+1}\left[(n-r+1)(n-r)\ln\frac{n-r+1}{n-r}\right.$$

$$\left. - (n-r)\right] = \frac{(n-k+2)(n-k+1)(n-k)}{4(n+2)}\ln\frac{n-k+2}{n-k} - \frac{(n-k+1)(n-k)}{2(n+2)} - (n-k+1)$$

$$(n-k)\ln\frac{n-k+1}{n-k} + (n-k) - \sum_{r=k+1}^{n} \frac{r+2}{n+2} - \frac{\lambda}{2}(n+1) \qquad (2.7)$$

For $k = n$, (2.7) implies $\lambda = 0$ (the indeterminate form $-(n-k)\ln(n-k)$ is zero by use of (2.4) and (2.5)).

Using (2.5) in (2.6) for $k = n$ provides:

$$\frac{\partial\Phi}{\partial\Delta_n} = \iint_{0\le u\le x<\infty} (\Delta_n - e^{-x})e^{-x}ne^{-u}(1-e^{-u})^{n-1}\,du\,dx = 0 \qquad (2.8)$$

which imples:

$$C_n = \frac{n+1}{n+2} .$$

To simplify notation, let:

$$F_1(k) = (n-k+1)(n-k)\ln\frac{n-k+1}{n-k} - 2(n-k+2)(n-k+1) + 2(n-k+2)(n-k+1)^2$$

$$\ln\frac{n-k+2}{n-k+1}$$

$$F_2(k) = 2(n-k)(n-k+1) - (2n-2k+1)(n-k+1)(n-k)\ln\frac{n-k+1}{n-k}$$

$$F_3(k) = (n-k+1)(n-k)\ln\frac{n-k+1}{n-k} - (n-k+2)(n-k+1)\ln\frac{n-k+2}{n-k+1}$$

$$H(k) = \frac{(n-k+2)(n-k+1)(n-k)}{4(n+2)}\ln\frac{n-k+2}{n-k} - \frac{(n-k+1)(n-k)}{2(n+2)} - (n-k+1)(n-k)$$

$$\ln\frac{n-k+1}{n-k} + (n-k) - \sum_{r=k+1}^{n} \frac{r+1}{n+2} - \frac{\lambda(n+1)}{2}$$

Equation (2.7) may be written in matrix form:

$$\underline{A}\,\underline{C} = \underline{H} \tag{2.9}$$

where:

$$\underline{A} = \begin{vmatrix} F_1(1) & F_3(2) & F_3(3) & \cdots & F_3(n-1) & F_3(n) \\ F_2(1) & F_1(2) & F_3(3) & \cdots & F_3(n-1) & F_3(n) \\ 0 & F_2(2) & F_1(3) & \cdots & F_3(n-1) & F_3(n) \\ 0 & 0 & F_2(3) & \cdots & F_3(n-1) & F_3(n) \\ 0 & 0 & 0 & \cdots & F_3(n-1) & F_3(n) \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & F_3(n-1) & F_3(n) \\ 0 & 0 & 0 & \cdots & F_1(n-1) & F_3(n) \\ 0 & 0 & 0 & \cdots & F_2(n-1) & F_1(n) \end{vmatrix}$$

$$\underline{C} = \begin{vmatrix} C_1 \\ C_2 \\ \cdot \\ \cdot \\ \cdot \\ C_{n+1} \end{vmatrix} \quad \text{and} \quad \underline{H} = \begin{vmatrix} H(0) \\ H(1) \\ \cdot \\ \cdot \\ \cdot \\ H(n) \end{vmatrix}$$

Successive row subtraction with the last row and the last column discarded (since C_n has already been determined) results in $\underline{A}$ becoming a tri-diagonal matrix of the form:

$$\underline{A}' = \begin{vmatrix} G_1(1) & G_2(1) & 0 & 0 & \cdots & 0 & 0 & 0 \\ G_2(1) & G_1(2) & G_2(2) & 0 & \cdots & 0 & 0 & 0 \\ 0 & G_2(2) & G_1(3) & G_2(3) & \cdots & 0 & 0 & 0 \\ 0 & 0 & G_2(3) & G_1(4) & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & G_2(4) & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & G_1(n-3) & G_2(n-3) & 0 \\ 0 & 0 & 0 & 0 & \cdots & G_2(n-3) & G_1(n-2) & G_2(n-2) \\ 0 & 0 & 0 & 0 & \cdots & 0 & G_2(n-2) & G_1(n-1) \end{vmatrix}$$

where:

$$G_1(k) = F_1(k) - F_2(k)$$

$$G_2(k) = F_3(k) - F_1(k) = F_2(k-1)$$

and

$$H' = \begin{vmatrix} H'(1) \\ H'(2) \\ \vdots \\ H'(n-1) \end{vmatrix} \text{ where } H'(k) = H(k-1) - H(k).$$

Equation (2.9) becomes:

$$\underline{A}' \ \underline{C} = \underline{H}' \qquad (2.91)$$

Equation (2.91) represents a 2nd order linear difference equation with variable coefficients. An explicit solution appears to be out of reach, forcing the use of numerical methods.

This matrix equation can be solved by triangular decomposition for any given sample size. The algorithm for this deterministic solution and the associated Fortran subroutine, TRID, are described

in Ref. 2. Coefficients for various sample sizes were computed using TRID and are shown with the corresponding Pyke coefficients in Table I.

TABLE I. COEFFICIENTS OF THE CONTINUOUS ESTIMATORS

N	C	MINIMUM RISK	PYKE
1			
	1	0.6666667	0.5000000
2			
	1	0.3383387	0.3333333
	2	0.7500000	0.6666667
3			
	1	0.3432969	0.2500000
	2	0.4987318	0.5000000
	3	0.8000000	0.7500000
4			
	1	0.2800868	0.2000000
	2	0.4240374	0.4000000
	3	0.5887811	0.6000000
	4	0.8333333	0.8000000
5			
	1	0.2394752	0.1666667
	2	0.3578057	0.3333333
	3	0.5126778	0.5000000
	4	0.6460857	0.6666667
	5	0.8571429	0.8333333
6			
	1	0.2085332	0.1428571
	2	0.3125745	0.2857143
	3	0.4437924	0.4285714
	4	0.5721471	0.5714286
	5	0.6906526	0.7142857
	6	0.8750000	0.8571429
7			
	1	0.1848704	0.1250000
	2	0.2769643	0.2500000
	3	0.3940854	0.3750000
	4	0.5042591	0.5000000
	5	0.6200241	0.6250000
	6	0.7249480	0.7500000
	7	0.8888889	0.8750000
8			
	1	0.1660087	0.1111111
	2	0.2488631	0.2222222
	3	0.3539021	0.3333333
	4	0.4534583	0.4444444
	5	0.5541501	0.5555556
	6	0.6579414	0.6666667
	7	0.7524714	0.7777778
	8	0.9000000	0.8888889

N	C	MINIMUM RISK	PYKE
9			
	1	0.1506598	0.1000000
	2	0.2259314	0.2000000
	3	0.3213671	0.3000000
	4	0.4115223	0.4000000
	5	0.5034373	0.5000000
	6	0.5946056	0.6000000
	7	0.6890570	0.7000000
	8	0.7749696	0.8000000
	9	0.9090909	0.9000000
10			
	1	0.1379129	0.0909091
	2	0.2068969	0.1818182
	3	0.2943053	0.2727273
	4	0.3768931	0.3636364
	5	0.4608337	0.4545455
	6	0.5447461	0.5454545
	7	0.6284070	0.6363636
	8	0.7149642	0.7272727
	9	0.7937232	0.8181818
	10	0.9166667	0.9090909
20			
	1	0.0747509	0.0476190
	2	0.1123907	0.0952381
	3	0.1599652	0.1428571
	4	0.2048848	0.1904762
	5	0.2505187	0.2380952
	6	0.2959665	0.2857143
	7	0.3414700	0.3333333
	8	0.3869660	0.3809524
	9	0.4324732	0.4285714
	10	0.4779891	0.4761905
	11	0.5235179	0.5238095
	12	0.5690642	0.5714286
	13	0.6146325	0.6190476
	14	0.6602411	0.6666667
	15	0.7058782	0.7142857
	16	0.7516879	0.7619048
	17	0.7973108	0.8095238
	18	0.8445265	0.8571429
	19	0.8874853	0.9047619
	20	0.9545455	0.9523810

N	C	MINIMUM RISK	PYKE
50	1	0.0315038	0.0196078
	2	0.0474483	0.0392157
	3	0.0675612	0.0588235
	4	0.0865575	0.0784314
	5	0.1058530	0.0980392
	6	0.1250684	0.1176471
	7	0.1443054	0.1372549
	8	0.1635367	0.1568627
	9	0.1827696	0.1764706
	10	0.2020022	0.1960784
	11	0.2212349	0.2156863
	12	0.2404678	0.2352941
	13	0.2597008	0.2549020
	14	0.2789339	0.2745098
	15	0.2981671	0.2941176
	16	0.3174005	0.3137255
	17	0.3366341	0.3333333
	18	0.3558678	0.3529412
	19	0.3751017	0.3725490
	20	0.3943358	0.3921569
	21	0.4135701	0.4117647
	22	0.4328046	0.4313725
	23	0.4520395	0.4509804
	24	0.4712746	0.4705882
	25	0.4905101	0.4901961
	26	0.5097460	0.5098039
	27	0.5289823	0.5294118
	28	0.5482191	0.5490196
	29	0.5674565	0.5686275
	30	0.5866945	0.5882353
	31	0.6059332	0.6078431
	32	0.6251728	0.6274510
	33	0.6444134	0.6470588
	34	0.6636552	0.6666667
	35	0.6828984	0.6862745
	36	0.7021433	0.7058824
	37	0.7213902	0.7254902
	38	0.7406397	0.7450980
	39	0.7598924	0.7647059
	40	0.7791493	0.7843137
	41	0.7984114	0.8039216
	42	0.8176810	0.8235294
	43	0.8369599	0.8431373
	44	0.8562559	0.8627451
	45	0.8755638	0.8823529
	46	0.8949449	0.9019608
	47	0.9142469	0.9215686
	48	0.9342227	0.9411765
	49	0.9523976	0.9607843
	50	0.9807692	0.9803922

III. CALCULATION OF THE RISKS

From equation (2.2), the risk function for the minimum risk coefficients, R(min), is:

$$R(\min) = \sum_{r=0}^{n} \iiint\limits_{0\le u\le x\le v<\infty} \left[1 - e^{-x} - C_r - \frac{x-u}{v-u}\Delta_r\right]^2 e^{-x} f_{r,r+1}(u,v)\,du\,dv\,dx$$

$$= \sum_{r=0}^{n} \left\{ \frac{(1-C_r)^2}{n+1} + \frac{2(C_r-1)(n-r+1)}{(n+2)(n+1)} + \Delta_r \frac{(n-r+2)(n-r+1)(n-r)}{2(n+2)(n+1)} \right.$$

$$\left[\ln\frac{n-r+2}{n-r} - \frac{2}{n-r+2}\right] + 2\Delta_r(C_r-1)\,\frac{(n-r+1)(n-r)}{n+1}\left[\ln\frac{n-r+1}{n-r} - \right.$$

$$\left.\frac{1}{n-r+1}\right] + \Delta_r^2\,\frac{(n-r+1)(n-r)}{n+1}\left[\frac{2n-2r+1}{n-r+1} - 2(n-r)\ln\frac{n-r+1}{n-r}\right]$$

$$\left. + \frac{(n-r+1)(n-r+2)}{(n+3)(n+2)(n+1)} \right\} \tag{3.1}$$

The risk function for the Pyke estimator, R(Pyke), is:

$$R(\text{Pyke}) = \sum_{r=0}^{n} \iiint\limits_{0\le u\le x\le v<\infty} \left[1 - e^{-x} - \frac{r}{n+1} - \frac{x-u}{v-u}\left(\frac{1}{n+1}\right)\right]^2 e^{-x} f_{r,r+1}(u,v)\,du\,dv\,dx$$

$$= \sum_{r=0}^{n} \left\{ \frac{1}{(n+1)^3}\left[5(n-r)^2 + 5(n-r) + 1\right] - \frac{(n-r+1)(3n-3r+1)}{(n+1)^2(n+2)} \right.$$

$$+ \frac{(n-r+1)(n-r+2)}{(n+3)(n+2)(n+1)} - \frac{2(n-r)(n-r+1)(2n-2r+1)}{(n+1)^3}\ln\frac{n-r+1}{n-r}$$

$$\left. + \frac{(n-r)(n-r+1)(n-r+2)}{2(n+1)^2(n+2)}\ln\frac{n-r+2}{n-r} \right\} \tag{3.2}$$

The risk function for the sample distribution function, R(SDF), is:

$$R(SDF) = \sum_{r=0}^{n} \iiint\limits_{0 \le u \le x \le v < \infty} \left[1 - e^{-x} - \frac{r}{n}\right]^2 e^{-x} f_{r,r+1}(u,v)\,du\,dv\,dx$$

$$= \frac{1}{6n} \,. \qquad (3.3)$$

IV. RESULTS AND CONCLUSIONS

Values of each type risk were computed for various sample sizes using (3.1), (3.2) and (3.3) and are shown in Table II and Figure 1 where it becomes apparent that the risk of the Pyke estimator is significantly closer to the optimum than is the risk of the sample distribution function. All the risks converge to zero at the rate 1/n. Thus, n times the risk converges to a constant (1/6). It is interesting to compare the risks as they approach this asympototote. This is done in Table III and Figure 2.

These results coupled with those of Ref. 4 suggest that, given the criteria of minimizing expected squared error, the Pyke estimator should be used in lieu of the sample distribution function, particularly if the underlying population is suspected to be either exponential or uniform.

TABLE II. VALUE OF RISK FUNCTIONS FOR SAMPLE SIZE N

N	MINIMUM RISK	PYKE RISK	SDF RISK
1	0.0480993	0.0682656	0.1666667
2	0.0363529	0.0394270	0.0833333
3	0.0276659	0.0298089	0.0555556
4	0.0233801	0.0246450	0.0416667
5	0.0203317	0.0212262	0.0333333
6	0.0180346	0.0187209	0.0277778
7	0.0162237	0.0167779	0.0238095
8	0.0147538	0.0152155	0.0208333
9	0.0135338	0.0139269	0.0185185
10	0.0125036	0.0128436	0.0166667
12	0.0108566	0.0111197	0.0138889
14	0.0095964	0.0098068	0.0119048
16	0.0086000	0.0087725	0.0104167
18	0.0077919	0.0079360	0.0092593
20	0.0071232	0.0072455	0.0083333
30	0.0049866	0.0050497	0.0055556
40	0.0038370	0.0038755	0.0041667
50	0.0031184	0.0031444	0.0033333
60	0.0026266	0.0026453	0.0027778
70	0.0022689	0.0022830	0.0023810
80	0.0019969	0.0020079	0.0020833
90	0.0017832	0.0017921	0.0018519
100	0.0016108	0.0016181	0.0016667

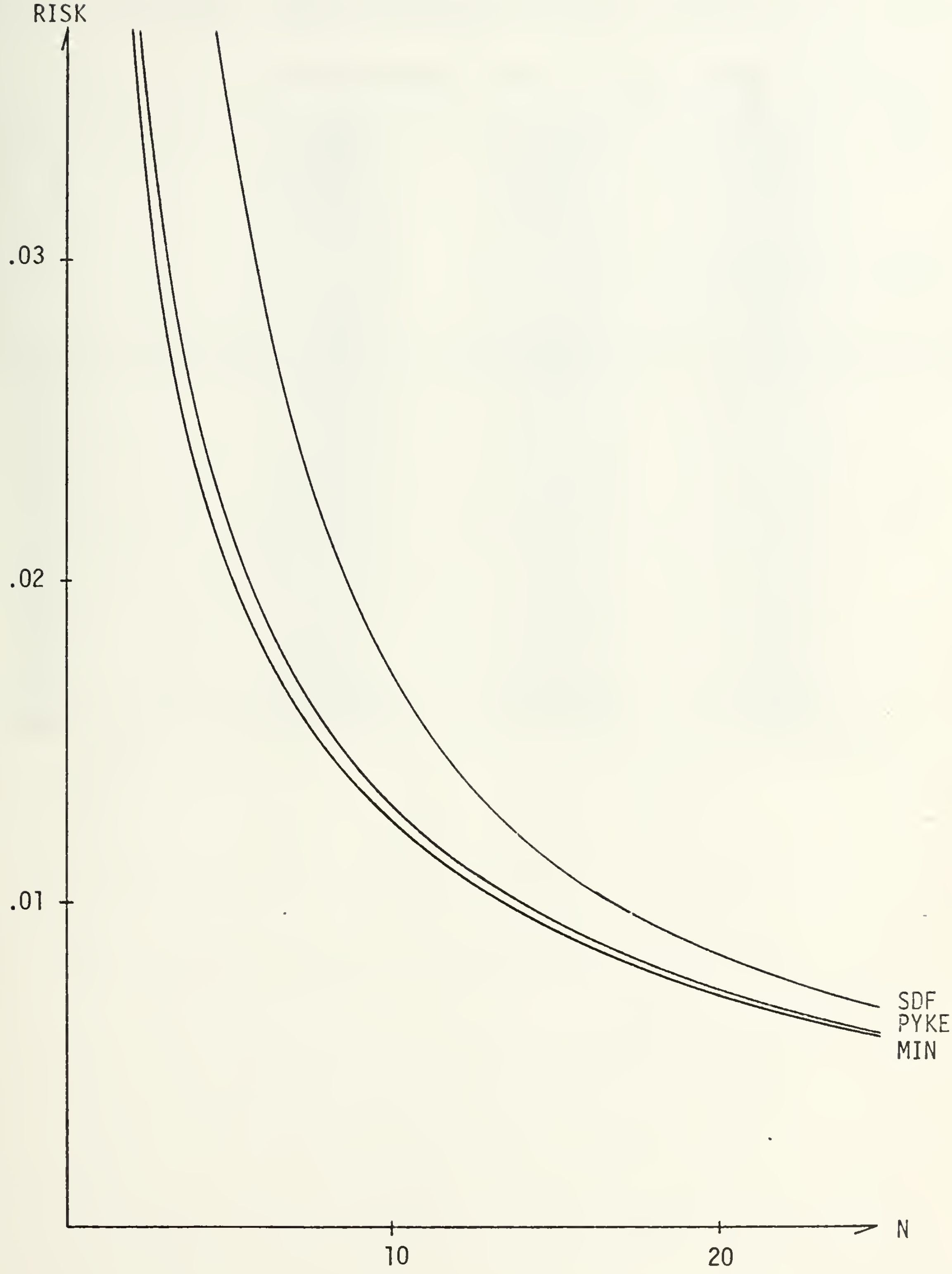

Figure 1.

TABLE III. VALUE OF THE RISK FUNCTIONS X SAMPLE SIZE

N	N(MINIMUM RISK)	N(PYKE RISK)	N(SDF RISK)
1	0.0480993	0.0682656	0.1666667
2	0.0727058	0.0788540	0.1666667
3.	0.0829976	0.0894267	0.1666667
4	0.0935202	0.0985800	0.1666667
5	0.1016586	0.1061310	0.1666667
6	0.1082079	0.1123254	0.1666667
7	0.1135658	0.1174451	0.1666667
8	0.1180301	0.1217241	0.1666667
9	0.1218044	0.1253423	0.1666667
10	0.1250362	0.1284358	0.1666667
12	0.1302796	0.1334367	0.1666667
14	0.1343494	0.1372958	0.1666667
16	0.1375995	0.1403596	0.1666667
18	0.1402547	0.1428488	0.1666667
20	0.1424646	0.1449103	0.1666667
30	0.1495975	0.1514916	0.1666667
40	0.1534788	0.1550203	0.1666667
50	0.1559200	0.1572188	0.1666667
60	0.1575977	0.1587195	0.1666667
70	0.1588217	0.1598089	0.1666667
80	0.1597542	0.1606358	0.1666667
90	0.1604884	0.1612848	0.1666667
100	0.1610815	0.1618077	0.1666667

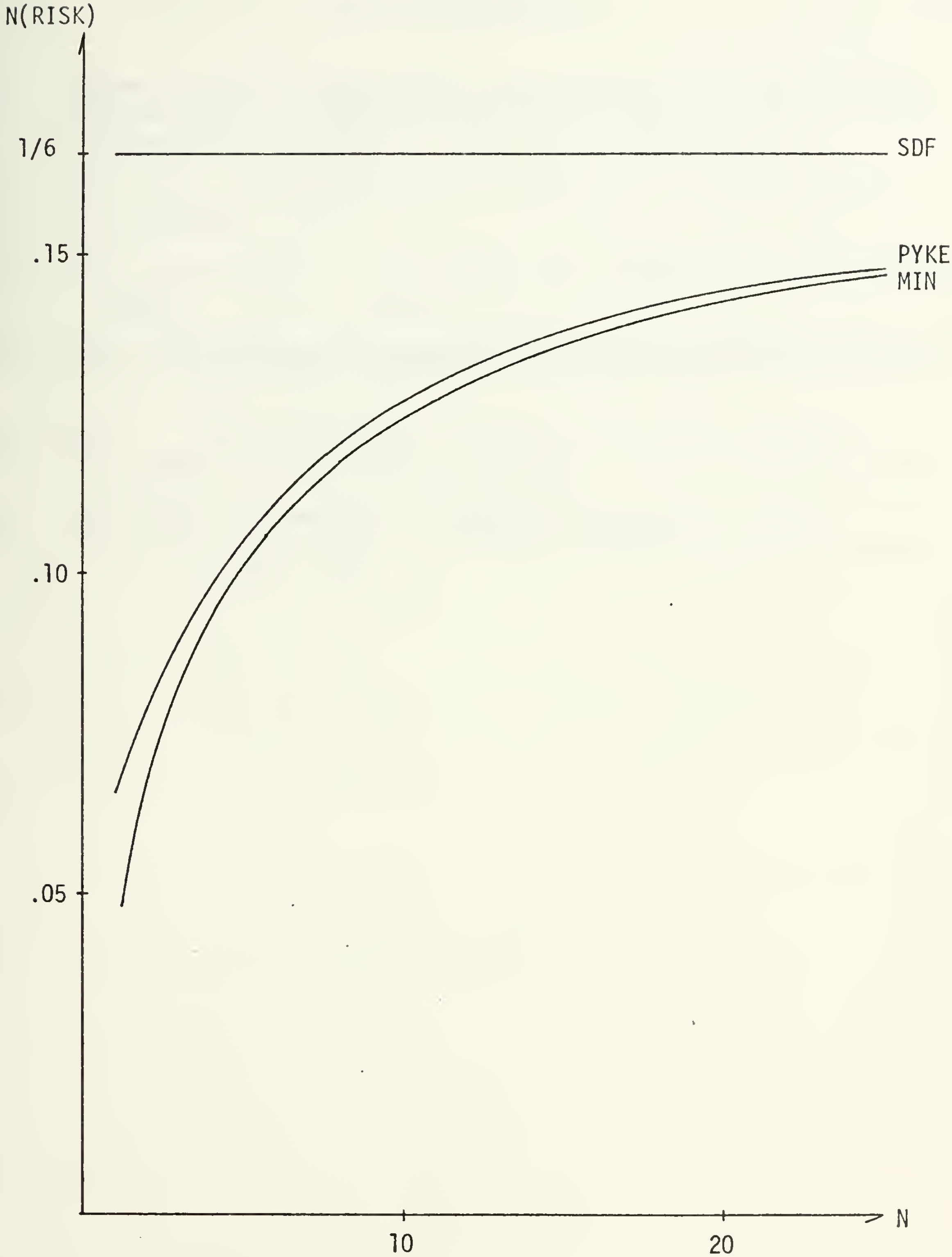

Figure 2.

LIST OF REFERENCES

1. Aggarwal, Om. P., "Some Minimax Invariant Procedures for Estimating a Cumulative Distribution Function", The Annals of Mathematical Statistics, v. 26, p. 450-463, 1955.

2. Conte, S.D., Elementary Numerical Analysis, p. 182-187, McGraw-Hill, 1965.

3. Hoel, P.G., Port, S.C., and Stone, C.J., Introduction to Statistical Theory, p. 166-171, Houghton-Mifflin, 1971.

4. Ozer, M., A Problem in Estimating a Distribution by Minimum Risk, Master Thesis, Naval Postgraduate School, 1970.

5. Pyke, R., "The Supremum and Infimum of the Poisson Process", The Annals of Mathematical Statistics, v. 30, p. 568-576, 1959.

6. Read, R.R., "The Asymptotic Inadmissability of the Sample Distribution Function", The Annals of Mathematical Statistics, v. 43, p. 89-95, 1970.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Chief of Naval Personnel Pers 11b Department of the Navy Washington, D.C. 20370	1
3.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
4.	Naval Postgraduate School Department of Operations Research and Administrative Sciences Monterey, California 93940	1
5.	Professor R.R. Read, Code 55 Re Department of Operations Research and Administrative Sciences Monterey, California 93940	1
6.	CPT Hugh Allan Kelley, USA CDEC Fort Ord, California 93941	1

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

[ORIGI]NATING ACTIVITY (Corporate author) [N]aval Postgraduate School [M]onterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified
	2b. GROUP

[REP]ORT TITLE

[T]he Pyke Estimator and the Exponential Distribution

[DE]SCRIPTIVE NOTES (Type of report and inclusive dates)

[M]aster's Thesis; (March 1973)

[AU]THOR(S) (First name, middle initial, last name)

[H]ugh Allan Kelley

[REP]ORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
[M]arch 1973	27	6

[C]ONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
[P]ROJECT NO.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

[D]ISTRIBUTION STATEMENT

Approved for public release; distribution unlimited

[S]UPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Naval Postgraduate School Monterey, California 93940

[A]BSTRACT

It is common to estimate a distribution by means of a step function. Such estimates can be made continuous by connecting the left points of the steps with straight line segments. In this paper, the best estimator of this class is found for data which is exponentially distributed using minimum risk. The risk is then compared with those of the sample distribution function and the Pyke estimator.

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
stimating Distribution						
xponential Distribution						
yke Estimator						
isk						
ample Distribution Function						

Thesis 143989
K265 Kelley
c.1 The Pyke estimator and
the exponential distri-
bution.